COVID-19 Case Surveillance Public Use Data Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value "Unknown" when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value "Missing" when jurisdictions do not provide a value, and the value "NA" when the value is suppressed as part of privacy protections.

Dataset version: 2021-08-17

Total records in dataset: 28,652,764

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	28,652,764	28,652,764
total_cells	343,833,168	85,958,292
missing_fields	69,924,595	964,293
missing_pct	20%	1%
complete_fields	273,908,573	84,993,999
complete_pct	80%	99%
unknown_fields	23,080,592	9,881,054
unknown_pct	7%	11%
suppressed_fields	118	118
suppressed_pct	0%	0%
available_fields	250,827,863	75,112,827
available_pct	73%	87%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	16	0.0%	684,368	2.4%
sex	25	0.0%	57,483	0.2%
age_group	77	0.0%	222,442	0.8%
records_with_any_field	77	0.0%	900,756	3.1%